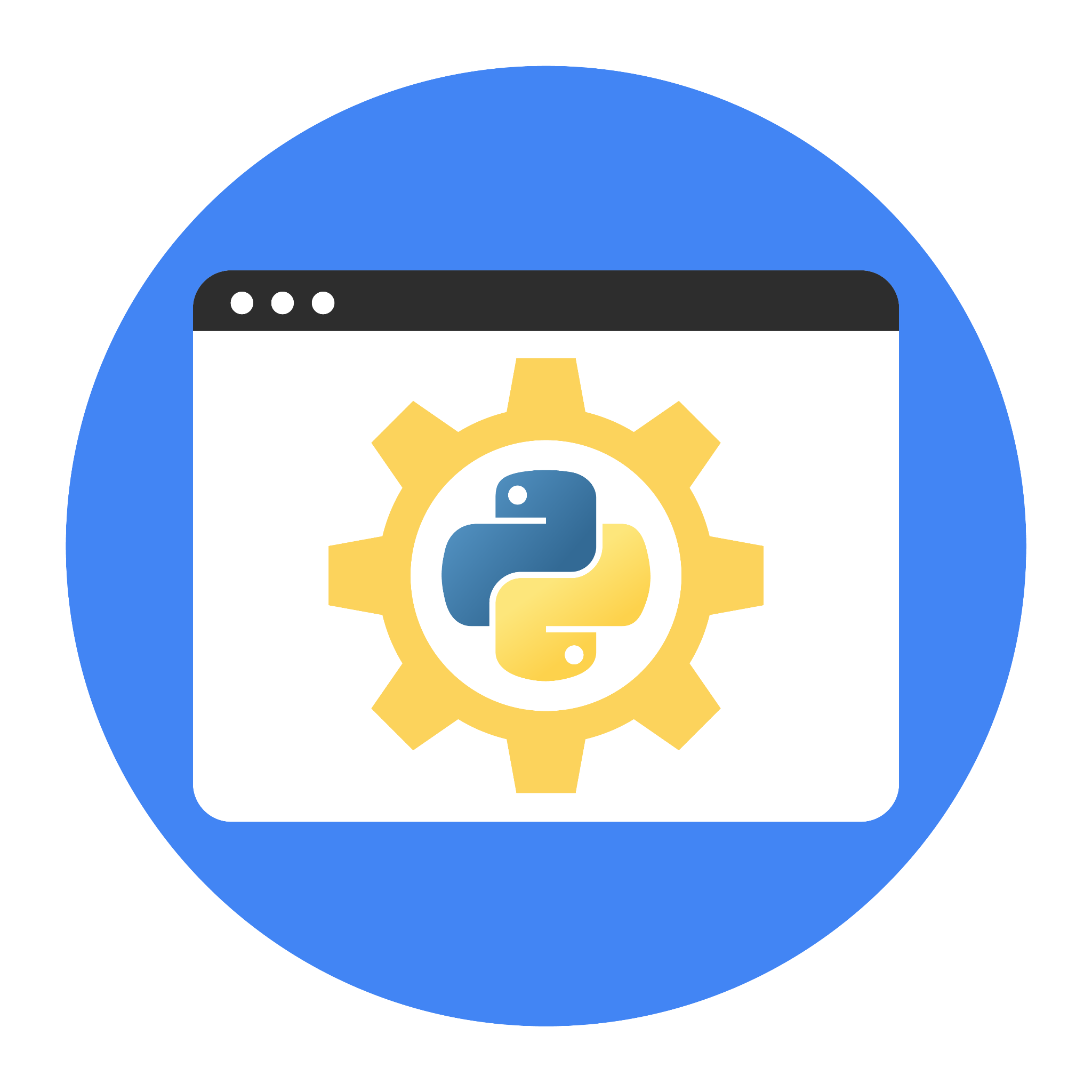
**Course Two**

# Get Started with Python



# Instructions

Use this PACE strategy document to record decisions and reflections as you work through this end-of-course project. You can use this document as a guide to consider your responses and reflections at different stages of the data analytical process. Additionally, the PACE strategy documents can be used as a resource when working on future projects.

# Course Project Recap

Regardless of which track you have chosen to complete, your goals for this project are:

* Complete the questions in the Course 2 PACE strategy document
* Answer the questions in the Jupyter notebook project file
* Complete coding prep work on project’s Jupyter notebook
* Summarize the column Dtypes
* Communicate important findings in the form of an executive summary

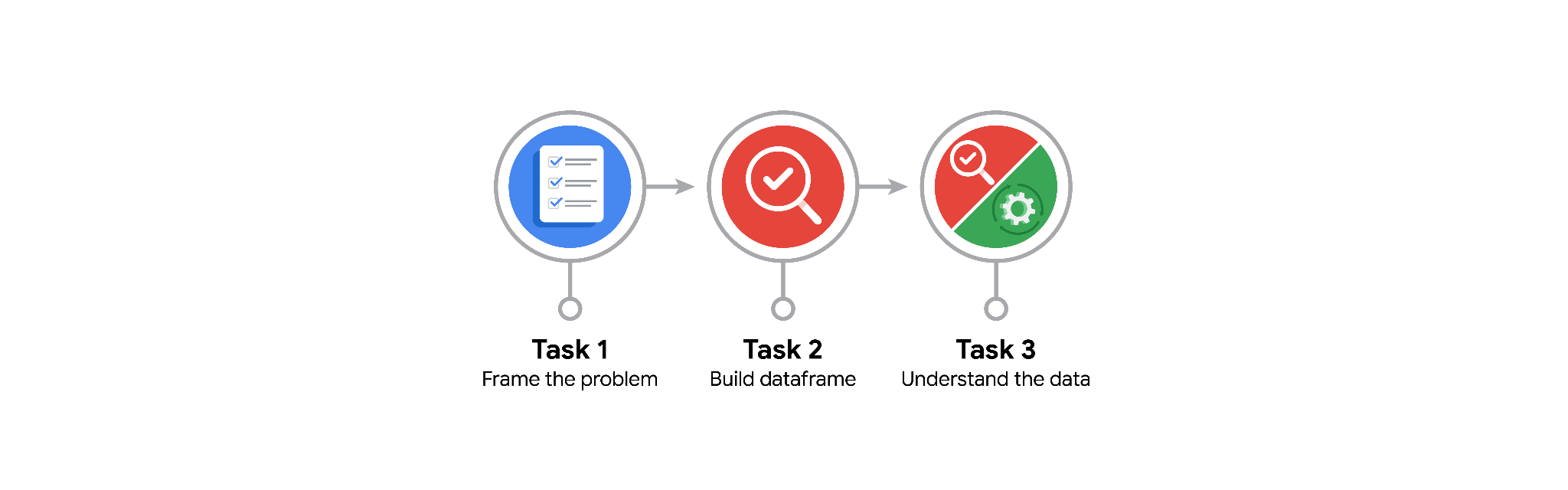
# Relevant Interview Questions

Completing the end-of-course project will help you respond these types of questions that are often asked during the interview process:

* Describe the steps you would take to clean and transform an unstructured data set.
* What specific things might you look for as part of your cleaning process?
* What are some of the outliers, anomalies, or unusual things you might look for in the data cleaning process that might impact analyses or ability to create insights?

**Reference Guide**

This project has three tasks; the visual below identifies how the stages of PACE are incorporated across those tasks.



**Data Project Questions & Considerations**

**PACE: Plan Stage**

* How can you best prepare to understand and organize the provided information?

Begin by exploring the dataset and study the columns of the dataset and it’s respective datatypes and perform summary statistics to develop a brief understanding of the dataset.

* What follow-along and self-review codebooks will help you perform this work?

Jupyter Notebook

* What are some additional activities a resourceful learner would perform before starting to code?

Plan the EDA process prior to the activity to follow a structured process.

**PACE: Analyze Stage**

* Will the available information be sufficient to achieve the goal based on your intuition and the analysis of the variables?

No, the given dataset provides information of users who are serious drivers rather than typical commuters, which is why more information is required to investigate further on the behavior of both these types of drivers.

* How would you build summary dataframe statistics and assess the min and max range of the data?

The summary dataframe statistics and the min and max range of the data can be accessed with the info() method.

* Do the averages of any of the data variables look unusual? Can you describe the interval data?

The average kilometers driven is unusually large for the number of drives which might indicate that the dataset includes serious drivers rather than typical commuters.

**PACE: Construct Stage**

**Note**: The Construct stage does not apply to this workflow. The PACE framework can be adapted to fit the specific requirements of any project.

**PACE: Execute Stage**

* Given your current knowledge of the data, what would you initially recommend to your manager to investigate further prior to performing exploratory data analysis?

Provide more data on the super-drivers to investigate further on their driving patterns and find out whether it is the reason for the increasing number of churns and whether these categories of drivers are the ones who churn the most.

* What data initially presents as containing anomalies?

Unusually high kilometers driven with a low number of days driven.

* What additional types of data could strengthen this dataset?

Users demographics, alternate navigation apps used, user feedback.